UNIT 12
STATISTICAL ANALYSIS PRESENTATION

As an assessment work for the Numerical Analysis Module

For my M. Sc. in Artificial Intelligence

Based on guidelines from the University of Essex

University of Essex

ASSIGNMENT TOPIC

“Alcohol is one of the leading issues in the UK. In 2017/18, there were 338 thousand estimated hospital admissions where the main reasons for admission to the hospital were attributable to alcohol and there were 5,843 alcohol-specific deaths. The number of deaths is 6% higher than in 2016 and an increase of 16% in 2007 (office for national statistics, 2019).”

DATA USED FOR THE TASK

Health Survey Data for England, 2011

SOFTWARE/TOOL USED

RSTUDIO 2022.02.3+492 "PRAIRIE TRILLIUM" RELEASE (1DB809B8323BA0A87C148D16EB84EFE39A8E7785, 2022-05-16) FOR MACOS

MOZILLA/5.0 (MACINTOSH; INTEL MAC OS X 12_5_1) APPLEWEBKIT/537.36 (KHTML, LIKE GECKO) QTWEBENGINE/5.12.10 CHROME/69.0.3497.128 SAFARI/537.36

CONTENTS

INTRODUCTION

Alcohol consumption in the UK has continued to rise compared to its neighbouring countries (EU), where the usage graph has either fallen or stabilised (*Heather, Nick*). This has impacted behaviour and mental health, leading to depression, anxiety and wellbeing. The isolation and restrictions during the COVID-19 pandemic played a significant factor in the increase in alcohol consumption *and particularly in young adults* (*Evans, Simon, et al)*.

Then why do we need to analyse?

This is required to find out how demographically and age-affecting is the consumption of alcohol that would help in building more focused efforts towards controlling the affects.

The data used for the following inferences are derived using

Health Survey Data for England, 2011.

References:

- Heather, Nick. "Britain's alcohol problem and what the UK government is (and is not) doing about it." Adicciones 18.3 (2006): 225-236.
- Evans, Simon, et al. "Effects of the COVID-19 lockdown on mental health, wellbeing, sleep, and alcohol use in a UK student sample." Psychiatry research 298 (2021): 113819.

DESCRIPTIVE STATISTICS

Total Sample	10,617

	Number	Percentage (out of total sample)
No. of People Drinks Alcohol	7,023	66.15
Men	3,247	30.58
Women	3,776	35.57
Highest Educational Level	1,749	16.47
Divorced	493	4.64
Separated	169	1.59
No. of People DOES NOT Drink Alcohol	3,594	33.85
Total Sample	10,617	

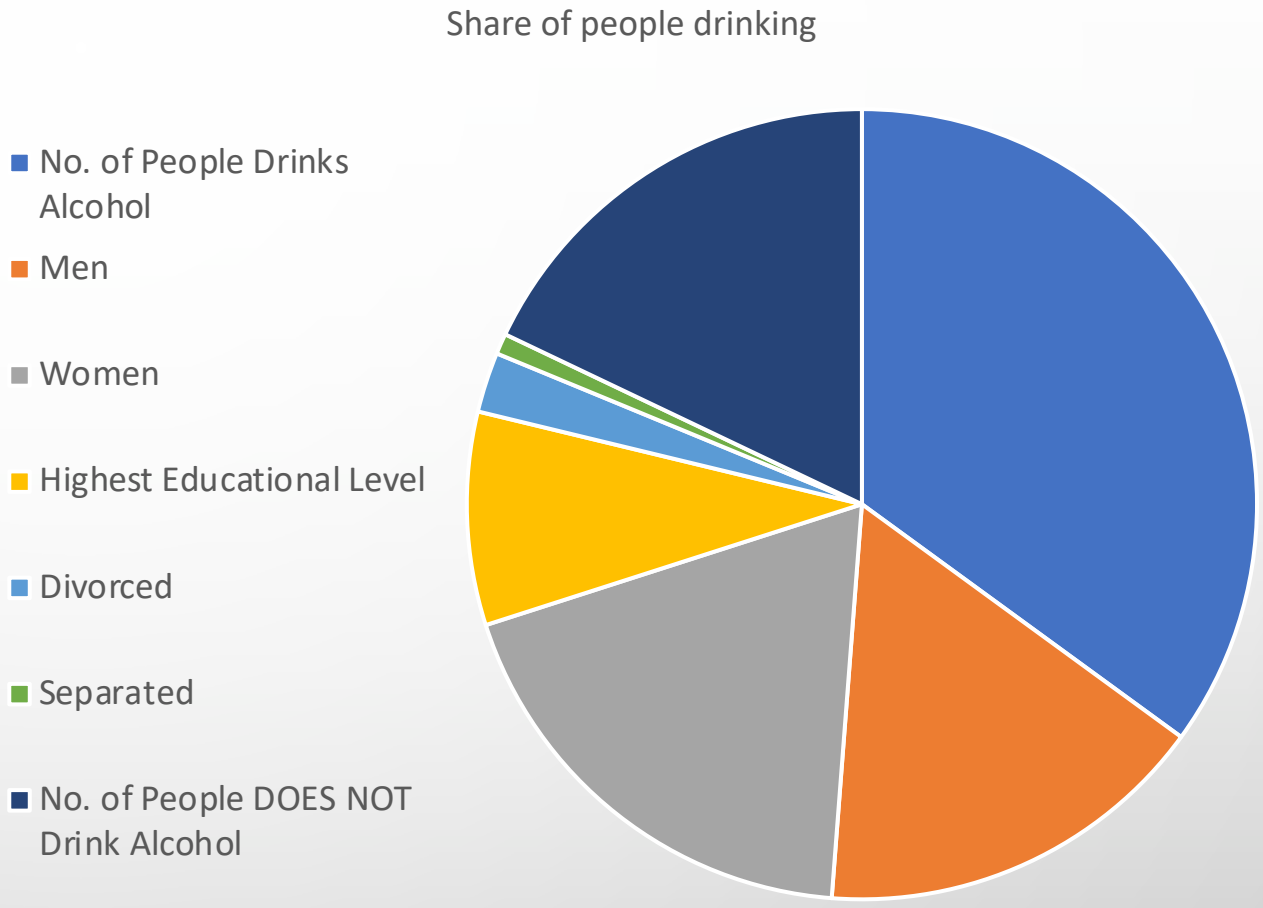
Based on the dataset and the analysis, it appears that there are more than half of the sample that has drinking Addiction. Between different sexes, they are quite comparable but based on the highest educational level, it shows that it is lower with higher education.

Refer the next slide that provides the various values affecting the age, BMI and household size

DESCRIPTIVE STATISTICS CONTD.

	Household Size	BMI	Age
Mean	3	27.41	49
Median	2	27	48
Mode	2	29	55
Minimum	1	13.56	16
Maximum	8	65.27	65
Range	1-8	13.56-65.27	16-96
Standard Deviation	1.2	5.26	17.68

Shows that an average household size of 3 within the middle-age sector is most affected.



INFERENTIAL STATISTICS

Is there a relation between gender and drinking status?

The chi-squared test was run between the Gender and the Drinking Status:

data: df$drinks and df$Sex

X-squared = 2.3797,

df = 1,

p-value = 0.1229

Since, the p-value is greater than 0.05, in that case we accept the null hypothesis that there is a relation between the genders and drinking status

Is there a relation between region and drinking status?

The chi-squared test was run between the Regions and the Drinking Status:

data: drinks_alcohol$gor1 and drinks_alcohol$totalwu

X-squared = 20967,

df = 20520,

p-value = 0.01415

Since, the p-value is lower than 0.05, in that case we reject the null hypothesis that there is a relation between the region and drinking

Is there a statistical difference between gender and height?

The t-test was run:

Welch Two Sample t-test

data: usages$Sex and usages$htval

t = -205.87, df = 10617, p-value < 1.82

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-131.4541 -128.9745

sample estimates:

mean of x mean of y

1.542997 131.757323

Is there a statistical difference between gender and weight?

The t-test was run:

Welch Two Sample t-test

data: usages$Sex and usages$wtval

t = -164.3, df = 10620, p-value < 0.04

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-56.04678 -54.72521

sample estimates:

mean of x mean of y

1.542997 56.928991

CONCLUSIONS AND RECOMMENDATIONS

Based on the various analysis and tests done on the dataset. It is inferred that there is an increase in the number of people drinking alcohol that does not relate to gender but more to age and educational level.

Heather, Nick. "Britain's alcohol problem and what the UK government is (and is not) doing about it." Adicciones 18.3 (2006): 225-236.

The recommendation is to review further samples and now adding more post Covid data along with social isolated effects and depressions on the population.

Oldham, M., Garnett, C., Brown, J., Kale, D., Shahab, L., & Herbec, A. (2021). Characterising the patterns of and factors associated with increased alcohol consumption since COVID-19 in a UK sample. Drug and alcohol review, 40(6), 890-899.

In conclusion and based on the current analysis, it is suggested that there should be more awareness and educational upliftment required to bring in the control for alcohol use.

MISCELLANEOUS

Subsequent slides supporting

The previous slides

Based on “R” calculations

R-SCRIPTS

The entire script is embedded in a word file →

RStudio

2011HSE_script.R

```
1  library(haven)
2
3  # read the data into a data table, X2011HSE
4  X2011HSE <- read_sav("/Users/gini/Google Drive/My Drive/01-Self/UoE/RProjects
5
6  # using package na.tools replace all NA with 0
7  library(na.tools)
8  X2011HSE[is_na(X2011HSE)]=0
9  print(X2011HSE)
10
11 # using the dplyr library for further features
12 library(dplyr)
13
14 # convert to local data frame for easy processing
15 alcohol_usages <- as_tibble(X2011HSE)
16
17 # select specific columns (sex, age, top qualification, total units, marital
18 usages <- select(alcohol_usages, Sex, Age, topqual3, totalwu, marstatc, HHSiz
19 print(usages)
20
21 # total sample
22 total_sample <- nrow(usages)
23 print(total_sample)
24
25 # percentage of people who drink alcohol
26
27 # 1. get data frame for people who drink alcohol
28 drinks_alcohol <- filter(select(usages,Sex, Age, topqual3, totalwu, marstatc,
29
30 print(drinks_alcohol)
```

Console | Terminal | Jobs

```
> library(haven)
>
> # read the data into a data table, X2011HSE
> X2011HSE <- read_sav("/Users/gini/Google Drive/My Drive/01-Self/UoE/RProjects/datafiles/HSE2011.sav")
>
> # using package na.tools replace all NA with 0
```

```
77  # -1      Not applicable
78  # 1       Single
79  # 2       Married
80  # 3       Civil partnership including spontaneous answers
81  # 4       Separated
82  # 5       Divorced
83  # 6       Widowed
84  # 7       Cohabitees
85  total_divorced_drink_alcohol <- nrow(filter(select(drinks_alcohol,totalwu, ma
86  print(total_divorced_drink_alcohol)
87
88  percent_divorced_drink_alcohol <- (total_divorced_drink_alcohol / total_sampl
89  print(percent_divorced_drink_alcohol)
90
91  total_separated_drink_alcohol <- nrow(filter(select(drinks_alcohol,totalwu, r
92  print(total_separated_drink_alcohol)
93
94  percent_separated_drink_alcohol <- (total_separated_drink_alcohol / total_sar
95  print(percent_separated_drink_alcohol)
96
97  # 11. mean, median, mode, min, max, range sd for household size, bmi, age
98
99  # using drinks alcohol data, filter out the 0 values
100 dataset <- filter(select(drinks_alcohol,Sex, Age, topqual3, totalwu, marstatc
101 dataset <- filter(select(dataset,Sex, Age, topqual3, totalwu, marstatc, HHSiz
102 dataset <- filter(select(dataset,Sex, Age, topqual3, totalwu, marstatc, HHSiz
103 print(dataset)
104
105
106 # mean
107 print(mean(dataset$HHSize))
108 print(mean(dataset$bmival))
109 print(mean(dataset$Age))
110
111 # median
```

```
> print(drinks_alcohol)
# A tibble: 7,023 × 10
        Sex   Age           topqual3 totalwu marstatc HHSize bmival    gor1 htval wtval
    <dbl+lbl> <dbl>          <dbl+lbl> <dbl+l> <dbl+lb>  <dbl> <dbl+> <dbl+l> <dbl> <dbl>
 1 2 [Female]    75 6 [Foreign/other]    0.058 5 [Divo…      1   25.3 6 [Eas…  162.  66.3
 2 2 [Female]    47 4 [NVQ2/GCE O Lev…   4.99  5 [Divo…      3    0   6 [Eas…    0    0
 3 1 [Male]      77 1 [NVQ4/NVQ5/Degr…  49.0   2 [Marr…      2   25.6 6 [Eas…  170.  74.2
 4 1 [Male]      44 3 [NVQ3/GCE A Lev…  30.2   1 [Sing…      1    0   6 [Eas…  168.   0
 5 1 [Male]      66 1 [NVQ4/NVQ5/Degr…  13.6   2 [Marr…      2    0   6 [Eas…    0    0
 6 1 [Male]      84 7 [No qualificati…  24.6   6 [Wido…      1    0   6 [Eas…    0    0
 7 1 [Male]      62 4 [NVQ2/GCE O Lev…   4.62  2 [Marr…      3    0   6 [Eas…    0    0
 8 1 [Male]      74 2 [Higher ed belo…  47.8   2 [Marr…      2    0   6 [Eas…    0    0
 9 2 [Female]    46 4 [NVQ2/GCE O Lev…   0.866 2 [Marr…      4   33.8 6 [Eas…  168.  95.7
10 1 [Male]      44 4 [NVQ2/GCE O Lev…   4.69  2 [Marr…      4   25.9 6 [Eas…  182.  85.4
# … with 7,013 more rows
>
> # 2. total number of people who drink alcohol
> total_drink_alcohol <- nrow(drinks_alcohol)
> print(total_drink_alcohol)
[1] 7023
>
> # 3. percent who drink alcohol
> percent_drink_alcohol <- (total_drink_alcohol / total_sample) * 100
> print(percent_drink_alcohol)
[1] 66.14863
>
> # 4. total number of males (1) who drink alcohol
> total_males_drink_alcohol <- nrow(filter(select(drinks_alcohol,totalwu, Sex),Sex==1))
> print(total_males_drink_alcohol)
[1] 3247
>
> # 5. percent of males who drink alcohol
> percent_males_drink_alcohol <- (total_males_drink_alcohol / total_sample) * 100
> print(percent_males_drink_alcohol)
[1] 30.58303
```